Evaluating Model Performance

It is not enough to simply train a model on some data and then assume that the model will subsequently perform well on future data. Instead, as we've mentioned previously, we need to split off a portion of our labeled data and reserve it for evaluating our model's final performance. We refer to this as the *test dataset*.

> *The **test dataset** is a portion of labeled data that is split off and reserved for model evaluation.*

If a model learns to perform well with the training data, but performs poorly with the test data, then there may be a problem that we will need to address before putting our model out into the real world. In practice, we will also need to decide what metrics we will use to evaluate performance, and whether there are any particular thresholds that the model needs to meet on these metrics in order for us to decide that it is "good enough."

When splitting the available data, it is important to preserve the *statistical properties* of that data. This means that the data in the training, validation, and test datasets need to have similar statistical properties as the original data to prevent bias in the trained model.

QUIZ QUESTION

A researcher has collected datasets from two neighboring cities in order to develop a model that predicts the sale price of a house based on various features. Which of the following approaches would be best for this researcher?

- Split the data by city, so that data for the first city is used to train the model, and data for the second city is used to evaluate the model.
- Use all of the data for training the model, as this will result in greater power and the best possible model optimization.
- ✓ Split the data randomly, so that some houses from each city are included in both the training dataset and the test dataset.

SUBMIT

NEXT